

THE SCORE OF CONDITIONALLY HETEROSKEDASTIC DYNAMIC REGRESSION MODELS WITH STUDENT t INNOVATIONS, AN LM TEST FOR MULTIVARIATE NORMALITY*

Gabriele Fiorentini, Enrique Sentana and Giorgio Calzolari**

WP-AD 2000-33

Correspondence to Gabriele Fiorentini, Universidad de Alicante, Dpto. Fundamentos de Análisis Económico, Campus de San Vicente del Raspeig, 03071 Alicante, E-mail: ga...@merlin.fae.ua.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

First Edition December 2000.

Depósito Legal: V-5167-2000

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

* Financial support from CICYT, IVIE and MURST through the project "Statistical Models for Time Series Analysis" is gratefully acknowledged.

** G. Fiorentini: University of Alicante, ISIS-JRC and Ivie; E. Sentana: CEMFI; G. Calzolari: University of Florence.

THE SCORE OF CONDITIONALLY HETEROSKEDASTIC
DYNAMIC REGRESSION MODELS
WITH STUDENT t INNOVATIONS,
AN LM TEST FOR MULTIVARIATE NORMALITY

Gabriele Fiorentini, Enrique Sentana and Giorgio Calzolari

ABSTRACT

We provide numerically reliable analytical expressions for the score of conditionally heteroskedastic dynamic regression models when the conditional distribution is multivariate t . We also derive one-sided and 2-sided LM tests for multivariate normality versus multivariate t based on the first two moments of the (squared) norm of the standardised innovations evaluated at the Gaussian quasi-ML estimators of the conditional mean and variance parameters. We reinterpret them as specification tests for multivariate excess kurtosis, and show that they have power against leptokurtic alternatives. Finally, we analyse UK stock returns, and confirm that their conditional distribution has fat tails.

Keywords: Kurtosis, Inequality Constraints, ARCH, Financial Returns.

JEL: C51, C52

1 Introduction

Many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering effects (see e.g. Bollerslev, Chou and Kroner (1992) for a survey). This has been long realised, and two main alternative inference approaches have been proposed. The first one uses a “robust” estimation method, such as the Gaussian quasi-maximum likelihood (ML) procedure advocated by Bollerslev and Wooldridge (1992), which remains consistent even if the assumption of conditional normality is violated. The second one, in contrast, specifies a parametric leptokurtic distribution for the standardised innovations, such as the Student t distribution employed by Bollerslev (1987). While the second procedure will often yield more efficient estimators than the first if the assumed conditional distribution is correct, it has the disadvantage that it may end up sacrificing consistency when it is not (see Newey and Steigerwald (1997)).

Notwithstanding such considerations, a significant advantage of the quasi-ML approach in Bollerslev and Wooldridge (1992) is that they derived convenient closed-form expressions for the Gaussian log-likelihood score, which can be used to obtain numerically accurate extrema of the objective function. In contrast, estimation under an alternative distribution typically relies on numerical approximations to the derivatives, which are often poor. One of the objectives of our paper is to partly close the gap between the two approaches, by providing numerically reliable analytical expression for the score of the multivariate conditionally heteroskedastic dynamic regression model considered by Bollerslev and Wooldridge (1992), when the distribution of the innovations is assumed to be proportional to a multivariate t . As is well known, the t distribution nests the normal as a limiting case, but has generally fatter tails. As documented by McCullough and Vinod (1999), the use of analytical derivatives in the estimation routine should consid-

erably improve the numerical accuracy of the resulting estimates. This should be particularly true in our case, because even with fairly large sample sizes, it becomes very difficult to numerically distinguish a standardised t with 100 degrees of freedom from another one with 5,000 degrees of freedom, or indeed from their Gaussian limit.¹

In addition, we derive a Lagrange Multiplier (LM) test for the null hypothesis of multivariate normality versus the alternative of multivariate t , whose one-sided version is asymptotically equivalent to the corresponding Likelihood Ratio (LR) and Wald tests. As usual, the main advantage of the LM test is that it is extremely simple to implement, because it only requires estimates of the standardised innovations evaluated at precisely the Gaussian quasi-ML estimators of the conditional mean and variance parameters. We also re-interpret our proposed test as a moment specification test of multivariate excess kurtosis, and show that it has non-trivial power against leptokurtic multivariate distributions. Therefore, it is not surprising that for some popular models, our proposed test coincides with the kurtosis component of Mardia's (1970) test for multivariate normality, which in turn reduces to the well-known Jarque and Bera (1980) test in the univariate case. However, it is important to stress that those tests cannot be directly applied to standardised innovations in more general models, e.g. when the innovations are conditionally heteroskedastic. Finally, we include an illustrative empirical application to UK stock returns, which confirms that their conditional distribution has rather fat tails.

The rest of the paper is organised as follows. First, we obtain closed-form

¹For instance, Microfit 4.0, which uses numerical derivatives to compute the score of univariate conditionally heteroskedastic regression models with standardised Student t innovations, explicitly warns the user that numerical accuracy cannot be achieved when the estimated degrees of freedom parameter is larger than 25, and recommends the use of the normal distribution instead (see Pesaran and Pesaran (1997), p. 457).

expressions for the log-likelihood score vector in section 2. Then, in section 3, we introduce our proposed LM test, discuss its properties, relate it to the existing literature, and present the empirical results. Finally, our conclusions can be found in section 4. Proofs are gathered in the appendix.

2 Analytical derivatives

2.1 A multivariate conditionally heteroskedastic dynamic regression model with Student t innovations

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N dependent variables, y_t , is typically assumed to be generated by the following equations:

$$\begin{aligned} y_t &= \mu_{0t} + S_{0t}^{1=2} \epsilon_t \\ \mu_{0t} &= \mu_0(x_t) \\ S_{0t} &= S(x_t) \end{aligned}$$

where $\mu_0(\cdot)$ and $\text{vech}[S(\cdot)]$ are $N \times 1$ and $N(N+1)/2 \times 1$ vector of functions known up to the $p \times 1$ vector of true parameter values μ_0 , x_t are k predetermined explanatory variables, which may contain contemporaneous conditioning variables z_t , as well as past values of y_t and z_t , I_{t-1} denotes the information set available at $t-1$, $S_{0t}^{1=2}$ is an $N \times N$ "square root" matrix such that $S_{0t}^{1=2} S_{0t}^{1=2\top} = S_{0t}$, and ϵ_t is a vector martingale difference sequence satisfying $E(\epsilon_t | z_t; I_{t-1}) = 0$ and $V(\epsilon_t | z_t; I_{t-1}) = I_N$. As a consequence,

$$\begin{aligned} E(y_t | z_t; I_{t-1}; \mu_0) &= \mu_{0t} \\ V(y_t | z_t; I_{t-1}; \mu_0) &= S_{0t} \end{aligned}$$

As in Bollerslev (1987), Baillie and Bollerslev (1989) and Harvey, Ruiz and Sentana (1992) among many others, our approach is based on the t-distribution. In particular, we assume that conditional on z_t and $I_{t|t-1}$, u_t^a is independent and identically distributed as a standardised multivariate t with ν_0 degrees of freedom. That is,

$$u_t^a = \frac{S}{\nu_0 - 2} \frac{u_t^y}{\nu_t},$$

where u_t^y is a multivariate standard normal variate, and ν_t an independent \hat{A}^2 random variable with $\nu_0 > 2$ degrees of freedom.² As is well known, the multivariate Student t approaches the multivariate normal as $\nu_0 \rightarrow \infty$, but has generally fatter tails (see e.g. Zellner (1971)). For that reason, it is often more convenient to use the reciprocal of the degrees of freedom parameter, $\gamma_0 = 1/\nu_0$, as a measure of tail thickness, which will always remain in the finite range $0 \leq \gamma_0 < 1/2$ under our assumptions.³

2.2 The log-likelihood function

Let $\hat{A} = (\mu^0; \gamma)^0$ denote the $p + 1$ parameters of interest. The log-likelihood function of a sample of size T (ignoring initial conditions)⁴ takes the form $L_T(\hat{A}) = \sum_{t=1}^T l_t(\hat{A})$, with $l_t(\hat{A}) = c(\gamma) + g_t(\hat{A})$:

²For the degrees of freedom to take any real value above 2, ν_t must in fact be an independent Gamma variate with mean ν_0 and variance $2\nu_0$.

³Note that for $\gamma_0 \geq 0.5$ the standardised t-distribution cannot be defined because the variance of the non-standardised t-distribution becomes infinite when $\nu_0 \leq 2$.

⁴Nevertheless, it is important to stress that since both $\hat{h}^1(x_t; \mu)$ and $\hat{S}(x_t; \mu)$ are often recursively defined (as in e.g. Arma or Garch models), it may be necessary to choose some initial values to start up the recursions. As pointed out by Fiorentini, Calzolari and Panatoni (1996), this fact should be taken into account in computing the score analytically, in order to make the results exactly comparable with those obtained by using numerical derivatives.

$$c(\gamma) = \ln \Gamma\left(\frac{N}{2} + \frac{1}{2\gamma}\right) - \ln \Gamma\left(\frac{N}{2}\right) - \frac{N}{2} \ln \frac{1}{\gamma} - \frac{N}{2} \ln \frac{1}{2} \quad (1)$$

and

$$g_t(\hat{A}) = -\frac{1}{2} \ln |S_t(\mu)| - \frac{N}{2} + \frac{1}{2\gamma} \log \left(1 + \frac{\gamma}{(1-2\gamma)} \kappa_t(\mu)\right) \quad (2)$$

where $\Gamma(\cdot)$ is Euler's gamma (or factorial) function, $S_t(\mu) = S(x_t; \mu)$, $\kappa_t(\mu) = \sigma_t^2(\mu) \times \sigma_t^2(\mu)$, $\sigma_t^2(\mu) = S_t^{1-2}(\mu) \sigma_t^2(\mu)$, $\sigma_t(\mu) = y_t - \sigma_t(\mu)$, and $\sigma_t(\mu) = \sigma(x_t; \mu)$. Not surprisingly, it can be readily verified that

$$\begin{aligned} \lim_{\gamma \rightarrow 0^+} c(\gamma) &= -\frac{N}{2} \ln 2 \\ \lim_{\gamma \rightarrow 0^+} g_t(\hat{A}) &= -\frac{1}{2} \ln |S_t(\mu)| - \frac{1}{2} \kappa_t(\mu) \end{aligned}$$

which confirms that $L_T(\mu; 0)$ collapses to a conditionally Gaussian log-likelihood.

Given the nonlinear nature of the model, a numerical optimisation procedure is usually required to obtain ML estimates of \hat{A} . Assuming that all the elements of $\sigma(x_t; \mu)$ and $S(x_t; \mu)$ are differentiable functions of μ , we can use a standard gradient method, where the required derivatives can be numerically approximated by re-evaluating $L_T(\hat{A})$ with each parameter in turn shifted by a small amount. But as we shall show in the next subsection, in this case it is also possible to obtain simple analytical expressions for the score.

The use of analytical derivatives in the estimation routine, as opposed to their numerical counterparts, should considerably improve the accuracy of the resulting estimates (see McCullough and Vinod (1999)). This is particularly true in our case, because even if the sample size T is large, the Student's t -based log-likelihood function is often rather flat for very small values of γ , and it becomes very difficult to numerically distinguish a standardised t with 100 degrees of freedom from another one with 5,000 degrees of freedom, or indeed from their Gaussian limit.

The analytical derivatives that we shall obtain could also be used even if the coefficients of the model were reparametrised as $\hat{A} = f(\frac{1}{2})$, with $\frac{1}{2}$ unconstrained, in order to maximise the unrestricted log-likelihood function $L_T[f(\frac{1}{2})] = \mathcal{L}_T(\frac{1}{2})$. In particular, we would have that

$$\frac{\partial \mathcal{L}_T(\frac{1}{2})}{\partial \frac{1}{2}} = \frac{\partial L_T(\hat{A})}{\partial \hat{A}} \frac{\partial f(\frac{1}{2})}{\partial \frac{1}{2}} \quad (3)$$

Nevertheless, one has to be careful with such transformations, because they may e.g. introduce false extrema (see section 7.4 of Gill, Murray and Wright (1981)).

2.3 The score function

Let $s_t(\hat{A}) = \partial l_t(\hat{A}) / \partial \hat{A}$ denote the score function, and partition it in two blocks:

$$\begin{aligned} s_{\mu t}(\hat{A}) &= \frac{\partial l_t(\hat{A})}{\partial \mu} = \frac{\partial g_t(\hat{A})}{\partial \hat{A}} \\ s_{\gamma t}(\hat{A}) &= \frac{\partial l_t(\hat{A}) / \partial \gamma}{\partial \gamma} = \frac{\partial c(\gamma)}{\partial \gamma} + \frac{\partial g_t(\hat{A})}{\partial \gamma} \end{aligned}$$

After tedious but otherwise straightforward algebraic manipulations, we can show that:

$$\begin{aligned} s_{\mu t}(\hat{A}) &= \frac{\partial^1 l_t(\mu)}{\partial \mu} S_t^{i=2}(\mu) \frac{N' + 1}{1 - 2\gamma' + \gamma' \&_t(\mu)} \eta_t^{\alpha}(\mu) \\ &+ \frac{1}{2} \frac{\partial \text{vec}^0[S_t(\mu)]}{\partial \mu}^h S_t^{i=2}(\mu) \otimes S_t^{i=2}(\mu) \text{vec} \frac{N' + 1}{1 - 2\gamma' + \gamma' \&_t(\mu)} \eta_t^{\alpha}(\mu) \eta_t^{\alpha 0}(\mu) - I_N \end{aligned} \quad (4)$$

where the Jacobian matrices $\partial^1 l_t(\mu) / \partial \mu^0$ and $\partial \text{vec}[S_t(\mu)] / \partial \mu^0$ depend on the particular specification adopted.⁵

Notice that $s_{\mu t}(\mu; 0)$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992). But even if $\gamma_0 > 0$, we can prove directly that the score vector $s_{\mu t}(\hat{A})$ evaluated at the true parameter values has the martingale difference property (cf. Crowder (1976)). Specifically,

⁵See e.g. Sentana (2000) for the case of a conditionally heteroskedastic in mean factor model.

Proposition 1

$$E [s_{\mu t}(\hat{A}_0)|z_t; I_{t|t-1}; \hat{A}_0] = 0$$

Unlike in the Gaussian case, though, this result is no longer generally valid when the conditional distribution is misspecified (see also Newey and Steigerwald (1997)).

Similarly, we can show that for $\epsilon > 0$

$$\frac{\partial c(\epsilon)}{\partial \epsilon} = \frac{1}{2^{\epsilon/2}} \left[\tilde{A} \left(\frac{\mu}{2^{\epsilon}} \right) - \tilde{A} \left(\frac{N}{2} + \frac{1}{2^{\epsilon}} \right) + \frac{N^{\epsilon/2}}{(1-2^{\epsilon})} \right] \quad (5)$$

and

$$\frac{\partial g_t(\hat{A})}{\partial \epsilon} = \frac{1}{2^{\epsilon/2}} \ln \left(1 + \frac{\epsilon}{1-2^{\epsilon}} \right) \psi_t(\mu) - \frac{\mu}{2^{\epsilon}} + \frac{N}{2} \frac{\psi_t(\mu)}{(1-2^{\epsilon})[1-2^{\epsilon} + \epsilon \psi_t(\mu)]} \quad (6)$$

where $\tilde{A}(x) = \psi \ln \Gamma(x) = \psi x$ is the so-called di-gamma function (or Gauss' psi function; see Abramowitz and Stegun (1964)), which can be computed using standard routines. If we then take limits as $\epsilon \rightarrow 0$ from above, we can show that

$$\lim_{\epsilon \downarrow 0^+} \frac{\partial c(\epsilon)}{\partial \epsilon} = \frac{N(N+2)}{4}$$

and

$$\lim_{\epsilon \downarrow 0^+} \frac{\partial g_t(\hat{A})}{\partial \epsilon} = -\frac{N+2}{2} \psi_t(\mu) + \frac{1}{4} \psi_t^2(\mu)$$

so that

$$s_{\cdot t}(\mu; 0) = \frac{N(N+2)}{4} - \frac{N+2}{2} \psi_t(\mu) + \frac{1}{4} \psi_t^2(\mu) \quad (7)$$

where $s_{\cdot t}(\mu; 0)$ should be understood as a directional derivative.

Unfortunately, both $\partial g_t(\hat{A})/\partial \epsilon$ and especially $\partial c(\epsilon)/\partial \epsilon$ are numerically unstable for ϵ small. When $N = 1$, for instance, Figure 1 shows that the numerical accuracy in the computation of (5) is very poor for ϵ small enough, and eventually

breaks down. The implicit threshold value, τ say, is clearly hardware and software dependent, but in our experience, a value of τ equal to 10^{-4} can be regarded as safe.

When $0 \leq \epsilon < \tau$, we suggest to evaluate (5) and (6) by means of (directional) first order Taylor expansions around $\epsilon = 0$. Let us start with the first term. Straightforward manipulations show that for $\epsilon > 0$

$$\begin{aligned} \frac{\partial^2 c(\epsilon)}{\partial \epsilon^2} &= \frac{N}{2} \frac{4\epsilon - 1}{\epsilon^2 (1 - 2\epsilon)^2} - \frac{1}{\epsilon^3} \tilde{A} \left(\frac{1}{2\epsilon} \right) - \tilde{A} \left(\frac{N}{2} + \frac{1}{2\epsilon} \right) \\ &\quad + \frac{1}{4\epsilon^4} \tilde{A}^{(0)} \left(\frac{1}{2} N + \frac{1}{2\epsilon} \right) - \tilde{A}^{(0)} \left(\frac{1}{2\epsilon} \right), \end{aligned}$$

where $\tilde{A}^{(0)}(x) = \partial^2 \ln j(x) / \partial x^2$ is the so-called tri-gamma function (see Abramowitz and Stegun (1964)). Although $\partial^2 c(\epsilon) / \partial \epsilon^2$ is also rather unstable near the origin, if we again take limits as $\epsilon \rightarrow 0$ from above, we can show that

$$\lim_{\epsilon \rightarrow 0^+} \frac{\partial^2 c(\epsilon)}{\partial \epsilon^2} = -\frac{N(N+2)(N-5)}{6}$$

so that

$$\frac{\partial c(\epsilon)}{\partial \epsilon} = \frac{N(N+2)}{4} - \frac{N(N+2)(N-5)}{6} \epsilon + O(\epsilon^2) \quad (8)$$

Similarly, we obtain that

$$\begin{aligned} \frac{\partial g_t(\tilde{A})}{\partial \epsilon} &= -\frac{N+2}{2} \mathfrak{g}_t(\tilde{A}) + \frac{1}{4} \mathfrak{g}_t^2(\tilde{A}) \\ &\quad + -(4+2N) \mathfrak{g}_t(\tilde{A}) + \frac{N+4}{2} \mathfrak{g}_t^2(\tilde{A}) - \frac{1}{3} \mathfrak{g}_t^3(\tilde{A}) \epsilon + O(\epsilon^2) \end{aligned} \quad (9)$$

While Figure 1 confirms that (8) provides an excellent approximation for ϵ small, it is important to mention that (9) is only guaranteed to provide a good approximation if in addition $\mathfrak{g}_t(\tilde{A})$ is not excessively large.

3 An LM Test for Multivariate normality

3.1 The information matrix under the null

We can easily compute an LM (or efficient score) test for multivariate normality versus multivariate t distributed innovations on the basis of the value of the score of the log-likelihood function evaluated at the restricted parameter estimates $\hat{A}_T = \arg \max_A L_T(A)$ subject to $\gamma = 0$. Importantly, note that $\hat{A}_T = (\hat{\mu}_T^0; 0)^0$ is such that $\hat{\mu}_T$ are precisely the Gaussian quasi-ML estimators proposed by Bollerslev and Wooldridge (1992). Rather conveniently, it turns out that the information matrix is block-diagonal between μ and γ when $\gamma_0 = 0$, as the following Proposition shows:

Proposition 2 If $\gamma_0 = 0$, then

$$V [S_A(\mu_0; 0) | \hat{A}_0] = \begin{pmatrix} 2 & & \\ & 3 & \\ & 0 & 5 \end{pmatrix} \begin{pmatrix} V [S_{\mu t}(\mu_0; 0) | \hat{A}_0] & 0 \\ 0 & N(N+2)=2 \end{pmatrix}$$

where

$$V [S_{\mu t}(\mu_0; 0) | \hat{A}_0] = -E [h_{\mu \mu t}(\mu_0; 0) | \hat{A}_0] = E \left[\frac{1}{2} \frac{\partial^2 \log f_t(\mu_0)}{\partial \mu \partial \mu} S_t^{-1}(\mu_0) \frac{\partial \log f_t(\mu_0)}{\partial \mu} \right. \\ \left. + \frac{1}{2} \frac{\partial \text{vec}^0 [S_t(\mu_0)]}{\partial \mu} S_t^{-1}(\mu_0) \otimes S_t^{-1}(\mu_0) \frac{\partial \text{vec}^0 [S_t(\mu_0)]}{\partial \mu} \right] \hat{A}_0^{3/4}$$

As a result, the element of the inverse information matrix corresponding to the tail thickness parameter γ will be given by the reciprocal of the last diagonal element of the information matrix.

Note also that the block-diagonality of the information matrix implies that a joint LM test of multivariate normality and any other restrictions on the conditional mean and conditional variance parameters μ , can be decomposed in two additive components, the first of which would be precisely our proposed test (see Bera and McKenzie (1987)).

3.2 Two-sided tests

In view of Proposition 2, we can compute the information matrix version of the LM test simply as follows:

$$J_{2T}^I(\mu_T) = \frac{-T^{-1} \sum_{t=1}^T P_t^h N(N+2) \epsilon_t(\mu_T) + (1+4) \epsilon_t^2(\mu_T)}{N(N+2)} \quad (10)$$

which, importantly, only depends on the first two sample moments of $\epsilon_t(\mu_T)$.

If $H_0 : \gamma = 0$ is true, then $J_{2T}^I(\mu_T)$ will have an asymptotic chi-square distribution with one degree of freedom. The limiting distribution could be obtained directly as in Proposition 3 below by combining the block-diagonality of the information matrix with the following result:

Lemma 1 The squared Euclidean norm of the true standardised innovations, $\epsilon_t(\mu_0)$, is independently and identically distributed as a \hat{A}_N^2 random variable with N degrees of freedom under the null, and as $N(\rho_0 - 2) = \rho_0$ times an F variate with N and ρ_0 degrees of freedom under the alternative.

An asymptotically equivalent test, both under the null and under local alternatives, is given by the usual outer product version of the LM test:

$$J_{2T}^O(\mu_T) = \frac{-T^{-1} \sum_{t=1}^T P_t^h N(N+2) \epsilon_t(\mu_T) + (1+4) \epsilon_t^2(\mu_T)}{-T^{-1} \sum_{t=1}^T P_t^h N(N+2) \epsilon_t(\mu_T) + (1+4) \epsilon_t^2(\mu_T)} \quad (11)$$

which can be computed as T times the uncentred R^2 from the regression of 1 on $s_t(\mu_T; 0)$. Alternatively, we could use the Hessian version of the LM test, i.e.:

$$J_{2T}^H(\mu_T) = \frac{-T^{-1} \sum_{t=1}^T P_t^h N(N+2) \epsilon_t(\mu_T) + (1+4) \epsilon_t^2(\mu_T)}{-T^{-1} \sum_{t=1}^T P_t^h h_{\gamma\gamma}(\mu_T; 0)} \quad (12)$$

with

$$h_{\gamma\gamma}(\mu; 0) = -\frac{N(N+2)(N-5)}{6} - (4+2N)\epsilon_t(\mu) + \frac{N+4}{2}\epsilon_t^2(\mu) - \frac{1}{3}\epsilon_t^3(\mu)$$

where we have used the fact that

$$\lim_{\lambda \rightarrow 0^+} \frac{\partial^2 l_t(\hat{A})}{\partial \lambda^2} = \lim_{\lambda \rightarrow 0^+} \frac{\partial^2 c(\lambda)}{\partial \lambda^2} + \lim_{\lambda \rightarrow 0^+} \frac{\partial^2 g_t(\hat{A})}{\partial \lambda^2}$$

and have obtained the required expressions from (8) and (9).

Given that the numerators of $\hat{s}_{2T}^I(\mu_T)$, $\hat{s}_{2T}^O(\mu_T)$ and $\hat{s}_{2T}^H(\mu_T)$ coincide, while the denominators of $\hat{s}_{2T}^H(\mu_T)$ and $\hat{s}_{2T}^O(\mu_T)$ converge in probability to the denominator of $\hat{s}_{2T}^I(\mu_T)$, which contains no stochastic terms, we would expect a priori that $\hat{s}_{2T}^I(\mu_T)$ would be the version of the test with the smallest size distortions, followed by $\hat{s}_{2T}^H(\mu_T)$, whose denominator involves the first three sample moments of $\mathbf{x}_t(\mu)$, and finally $\hat{s}_{2T}^O(\mu_T)$, whose calculation also requires its fourth sample moment (see also Davidson and MacKinnon (1983)).

3.3 One-sided tests

It is important to mention that the fact that $\lambda = 0$ lies at the boundary of the admissible parameter space invalidates the usual \hat{A}^2 distribution of the LR and Wald tests, which under the null will be more concentrated towards the origin (see Andrews (2000) and the references therein, as well as the simulation evidence in Bollerslev (1987)). The intuition can be perhaps more easily obtained in terms of the Wald test. If λ could take both positive and negative values, then, under standard regularity conditions, its unrestricted ML estimator $\hat{\lambda}_T$ would have an approximately normal distribution with mean 0 and variance $2 = [TN(N + 2)]$ in large samples by virtue of Proposition 2. However, since $\hat{\lambda}_T$ cannot really be negative, then $\sqrt{T}\hat{\lambda}_T$ will in fact have an asymptotic normal distribution with mean 0 and variance $2 = [N(N + 2)]$ censored from below at 0. As a result, the Wald test will be an equally weighted mixture of a chi-square distribution with 0 degrees of freedom,⁶ and a chi-square distribution with 1 degree of freedom. In practice,

⁶By convention, \hat{A}_0^2 is a degenerate random variable that equals zero with probability 1.

obviously, we need simply compare the t-statistic $\sqrt{P} \frac{\bar{s}_T}{\sqrt{N(N+2)^{-1}}} \hat{\alpha}_T$ with the appropriate one-sided critical value from the normal tables. For analogous reasons, the asymptotic distribution of the LR test will also be degenerate half the time, and a chi-square with one degree of freedom the other half.

Although the above argument does not invalidate the distribution of the LM statistics (10), (11) and (12), intuition suggests that the one-sided nature of the alternative hypothesis should be taken into account to obtain a more powerful test (cf. Demos and Sentana (1998)). For that reason, we also propose a simple one-sided version of the LM test for multivariate normality. In particular, since

$$E[s_{\cdot t}(\mu_0; 0) | \hat{A}_0] > 0$$

when $\gamma_0 > 0$ in view of Proposition 3 below, we propose to use

$$s_{1T}^1(p_T) = s_{2T}^1(p_T) * \text{sign} \left(T^{1/2} \sum_t s_{\cdot t}(p_T; 0) \right)$$

as our one-sided LM test, and to compare it to the same 50:50 mixture of chi-squares 0 and 1. In this context, we would reject H_0 at the 100% significance level if the average score with respect to γ evaluated at the Gaussian quasi-ML estimators $\hat{A}_T = (\hat{p}_T; 0)$ is positive and $s_{1T}^1(p_T)$ exceeds the $100(1-2^{-\alpha})$ percentile of a \hat{A}_T^2 distribution. Since the Kuhn-Tucker (KT) multiplier associated with the inequality restriction $\gamma \geq 0$ is equal to $\max_{\gamma \geq 0} -T^{1/2} \sum_t s_{\cdot t}(p_T; 0)$, the one-sided LM test is asymptotically equivalent to the KT multiplier test introduced by Gourieroux, Holly and Monfort (1980), which in turn is equivalent in large samples to the LR and Wald tests. As we argued before, the reason is that those tests are implicitly one-sided in our context. In this respect, it is important to mention that in the case of a single restriction, those one-sided tests should be asymptotically locally more powerful (see e.g. Andrews (2000)).

Nevertheless, it is still interesting to compare the power properties of the one-sided and two-sided LM statistics. But given that the block-diagonality of the

information matrix is generally lost under the alternative of $\gamma_0 > 0$, and its exact form is unknown, we can only get closed form expressions for the case in which the standardised innovations u_t^α are directly observed.⁷ To do so, we shall use the following result:

Proposition 3 If u_t^α is independent and identically distributed as a standardised multivariate t random vector with $\nu_0 > 8$ degrees of freedom, then

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{s_t(\mu_0; 0) - E[s_t(\mu_0; 0) | \mu_0; \nu_0]}{\sqrt{E[s_t^2(\mu_0; 0) | \mu_0; \nu_0]}} \xrightarrow{d} N(0; 1)$$

where

$$\begin{aligned} E[s_t(\mu_0; 0) | \mu_0; \nu_0] &= \frac{N(N+2)}{4} \frac{\mu_0 - 2}{\nu_0 - 4} - 1 \\ E[s_t^2(\mu_0; 0) | \mu_0; \nu_0] &= -\frac{3N^2(N+2)^2}{16} + \frac{N(N+2)^2(3N+4)}{8} \frac{\nu_0 - 2}{\nu_0 - 4} \\ &\quad - \frac{N(N+2)^2(N+4)}{4} \frac{(\nu_0 - 2)^2}{(\nu_0 - 4)(\nu_0 - 6)} \\ &\quad + \frac{N(N+2)(N+4)(N+6)}{16} \frac{(\nu_0 - 2)^3}{(\nu_0 - 4)(\nu_0 - 6)(\nu_0 - 8)} \end{aligned}$$

On this basis, we can obtain the asymptotic power of the one-sided and two-sided variants of the information matrix version of the LM test for any possible significance level α . The results for $\alpha = .05$ are plotted in Figures 2, 3 and 4 for γ_0 in the range $0 \leq \gamma_0 \leq .04$, that is $\nu_0 \geq 25$. Not surprisingly, the power of both tests uniformly increases with the sample size T for a fixed alternative, and as we depart from the null for a given sample size. Importantly, their power also increases with the number of series N . As expected, the one-sided test is more

⁷In more realistic cases, though, the results are likely to be qualitatively similar.

powerful than the usual two-sided one.⁸ The difference is particularly noticeable for small departures from the null, which is precisely when power is generally low. For instance, when $\rho_0 = 100$, $T = 500$ and $N = 10$, the power of the one-sided test is almost 60% while the power of its two-sided counterpart is less than 50% (see Figure 4). Similarly, the one-sided tests for $N = 1$ and $N = 5$ are initially more powerful than the two-sided tests for $N = 2$ and $N = 10$ respectively.

3.4 Relationship with existing kurtosis tests

Following Mardia (1970), we can define the population coefficient of multivariate excess kurtosis as:

$$\kappa = \frac{E[\kappa_t^2(\mu_0)]}{N(N+2)} - 1; \quad (13)$$

which equals $2(\rho_0 - 4)$ for the multivariate t distribution, as well as its sample counterpart:

$$\kappa_T(\mu) = \frac{\sum_{t=1}^{T-1} P_T \kappa_t^2(\mu)}{N(N+2)} - 1 \quad (14)$$

On this basis, we can write the numerator of (10) as

$$\sum_{t=1}^{T-1} \kappa_t^2(\mu_T; 0) = \frac{N(N+2)}{4} \left(\sum_{t=1}^{T-1} \kappa_t^2(\mu_T) - 2 \sum_{t=1}^{T-1} \frac{\kappa_t(\mu_T)}{N} - 1 \right) \quad \#$$

Since κ is trivially 0 under the null from Lemma 1, our LM test of multivariate normality versus multivariate t is essentially a test of multivariate excess kurtosis. In fact, if we ignore the term

$$\sum_{t=1}^{T-1} \frac{\kappa_t(\mu_T)}{N} - 1 \quad \# \quad (15)$$

⁸However, as ρ_0 approaches 1=8 from below, the one-sided test loses power for fixed N and T , and eventually the two-sided test becomes more powerful. This is due to the fact that the variance of the score goes to infinity as $\rho_0 \rightarrow 8$ from Proposition 3.

(10) coincides with the kurtosis component of Mardia's (1970) test for multivariate normality, which in turn reduces to the popular Jarque and Bera (1980) test in the univariate case.

Since $T^{-1} \sum_{t=1}^T \mathbf{u}_t(\mathbf{p}_T) \mathbf{u}_t'(\mathbf{p}_T) = \mathbf{I}_N$ implies that (15) is identically 0, it follows from (4) that their tests are valid in nonlinear regression models with conditionally homoskedastic disturbances estimated by Gaussian quasi-ML, if the covariance matrix of the innovations, \mathbf{S}_t , is unrestricted and does not affect $\mathbf{u}_t(\mathbf{x}_t; \mu)$, and the conditional mean parameters and the elements of $\text{vech}(\mathbf{S})$ are variation free. However, ignoring (15) in more general contexts may lead to size distortions, because it is precisely the inclusion of such a term what makes $\mathbf{u}_t(\mu_0; 0)$ orthogonal to the other elements of the score. The same point was forcefully made by Davidson and MacKinnon (1993) in a univariate context (see section 16.7 of their textbook), and not surprisingly, their suggested test for excess kurtosis turns out to be equal to (11), the outer product version of our LM test. Similarly, the term (15) also appears explicitly in the Kiefer and Salmon (1983) LM test for univariate excess kurtosis based on a Hermite polynomial expansion of the density, which coincides in their context with the information matrix version of our test (10).

Several authors have recently suggested alternative multivariate generalisations of the Jarque-Bera test, which as far as kurtosis is concerned, consist in adding up the univariate kurtosis tests for each element of $\mathbf{u}_t(\mathbf{p}_T)$ (see e.g. Doornik and Hansen (1994), Lütkepohl (1993) or Kilian and Derimoglou (2000)). But apart from the issue discussed in the previous paragraph, a potential shortcoming of those tests is that they are not invariant to the way in which the residuals $\mathbf{u}_t(\mathbf{p}_T)$ are orthogonalised to obtain $\mathbf{u}_t^*(\mathbf{p}_T)$. For instance, while Doornik and Hansen (1994) obtain $\mathbf{S}_t^{1=2}(\mathbf{p}_T)$ from the spectral decomposition of $\mathbf{S}_t(\mathbf{p}_T)$, the other authors use a Cholesky decomposition. In this respect, note that by implicitly assuming that the excess kurtosis is the same for all possible linear

combinations of the true standardised innovations ϵ_t^a , we obtain a test statistic which is numerically invariant to orthogonal rotations of $S_t^{1=2}(\mu_T)$ (see also Mardia (1970)). If ϵ_t^a were directly observed, the relative power of the two testing procedures would depend on the exact nature of the alternative hypothesis. Given that the ϵ_{it}^a 's are independent across $i = 1; \dots; N$ under the null, the situation is completely analogous to the comparison between the one-sided tests for Arch(q) of Lee and King (1993) and Demos and Sentana (1998). In particular, if we define $\gamma_i = E(\epsilon_{it}^a) - 1$ for $i = 1; \dots; N$, our test would be more powerful against alternatives close to $\gamma_i = \gamma$ for all i , while the additive test would have more power when the γ_i 's were rather dispersed.

3.5 A re-interpretation of the LM test as a moment specification test

As usual, it is possible to re-interpret (10) as a moment specification test of the restriction

$$E[s_t(\mu_0; 0) | \bar{A}_0] = E\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\epsilon_t(\mu) + \frac{1}{4}\epsilon_t^2(\mu) - \bar{A}_0\right] = 0 \quad (16)$$

In order to analyse in which directions such a moment test has power, it is convenient to state the following auxiliary results, which correspond to Theorem 2.5 (iii), and Examples 2.4 and 2.5, respectively, in Fang, Kotz and Ng (1990):

Theorem 1 ϵ_t^\pm is distributed as a spherically symmetric multivariate random vector of dimension N if and only if $\epsilon_t^\pm = \rho_t u_t$, where u_t is uniformly distributed on the unit sphere surface in R^N , and ρ_t is a non-negative random variable which is independent of u_t .

Example 1 ϵ_t^y is distributed as a standardised multivariate normal random vector of dimension N if and only if $\epsilon_t^y = \rho_t u_t$, where u_t is uniformly distributed on the

unit sphere surface in \mathbb{R}^N , and χ_t^2 is an independent chi-square random variable with N degrees of freedom.

Example 2 \mathbf{u}_t^α is distributed as a standardised multivariate Student t random vector of dimension N if and only if $\mathbf{u}_t^\alpha = \frac{\rho}{(\nu_0 - 2)^{3/2}} \chi_t^2 \mathbf{u}_t$, where \mathbf{u}_t is uniformly distributed on the surface unit sphere in surface \mathbb{R}^N , and χ_t^2 and ν_t are two mutually independent chi-square random variables with N and ν_0 degrees of freedom respectively, independent of \mathbf{u}_t .

The variables \mathbf{u}_t and \mathbf{u}_t are usually referred to as the generating variate and the uniform base of the spherical distribution. In this light, our proposed LM test is simply a test of whether the (squared) generating variate $\mathbf{u}_t(\mu_0)$ is \hat{A}_N^2 against the alternative that it is proportional to an F_{N, ν_0} . But since our test is based on comparing the first two moments of $\mathbf{u}_t(\mu_0)$, the two-sided version has non-trivial power against any other spherically symmetric distribution for which $s_{\mathbf{u}_t}(\mu_0; 0)$ has expected value different from zero. For instance, if we consider the extreme case in which the true standardised disturbances were in fact uniformly distributed on the unit sphere surface in \mathbb{R}^N , so that $\mathbf{u}_t(\mu_0) = N \forall t$ and $\nu = -(N + 1) = (N + 2)$, then

$$s_{\mathbf{u}_t}(\mu_0; 0) = \frac{-N(N + 1)}{4};$$

which means that we would reject the null hypothesis with probability approaching one as T goes to infinite. On the other hand, the one-sided LM test only has power for the leptokurtic subclass of spherically symmetric distributions. Nevertheless, as we shall see in the next subsection, standardised residuals are frequently leptokurtic and rarely platykurtic in practice.

3.6 An empirical application to UK stock returns

In order to investigate the practical performance of the LM test for normality discussed in the previous subsections, we use the results in Sentana (1995) for monthly excess returns on UK stocks for the period 1971:2 to 1990:10 (237 observations). In particular, he estimated by Gaussian quasi-ML both a univariate $gqarch(1,1)$ -M model for the FT500 excess return series, and a conditionally heteroskedastic in mean latent factor model for the excess returns on 26 sectorial indices, with a $gqarch(1,1)$ parametrisation for the common factor, and a constant diagonal covariance matrix for the idiosyncratic terms. On the basis of the parameter estimates that he obtained, we generate the time series of (squared) Euclidean norms of the standardised innovations, $\epsilon_t(p_T)$. Then, we compute normal versions of the LM tests using the statistic:

$$\chi_T^1(p_T) = \sqrt{T} \frac{1}{T} \sum_t \frac{\sum_{i=1}^h [N(N+2) - (1+N) \epsilon_t(p_T) + \epsilon_t^2(p_T)]}{N(N+2)}$$

so that

$$\begin{aligned} \chi_{2T}^1(p_T) &= \chi_T^1(p_T) \\ \chi_{1T}^1(p_T) &= \chi_T^1(p_T) \cdot \text{sign} \left(\chi_T^1(p_T) \right) \end{aligned}$$

We find that $\chi_T^1(p_T)$ takes the value 13.71 for the aggregate stock market returns ($N = 1$), which is extremely significant regardless of whether we use a one-sided or a two-sided critical value. As expected, we also find that the corresponding test statistic is even higher (54.29) for sectorial returns ($N = 26$). As indicated in the introduction, such results are not very surprising in view of the existing empirical evidence, and simply reflect the fact that the standardised innovations are rather leptokurtic. In particular, the coefficient of multivariate excess kurtosis (14) for the sectorial data is 0.3698, which as the test statistic clearly

indicates, is very different from the theoretical value of $\gamma = 0$ under multivariate normality (see (13)).

In fact, we can easily obtain from $\hat{\gamma}_T(\hat{\mu}_T)$ a (two-stage) method of moments (MM) estimator of the degrees of freedom parameter ν that exploits the theoretical relationship $\gamma = 2/(\nu - 4)$, or $\nu = 4 + 2/\gamma$. The resulting MM estimate of ν is 9.41, which is rather close to the ML estimate of 9.73 obtained by Sentana (1991) when he considered a multivariate t distribution for the standardised innovations of the UK sectorial returns.

4 Conclusions

In the context of the general multivariate dynamic regression model with time-varying variances and covariances considered by Bollerslev and Wooldridge (1992), our two main contributions are:

1. We provide numerically reliable analytical expressions for the score vector when the distribution of the innovations is assumed to be proportional to a multivariate t .
2. We derive an LM test for multivariate normal versus multivariate t innovations, which is extremely simple to implement, because it is based on the first two sample moments of the (squared) Euclidean norm of the standardised innovations evaluated at the Gaussian quasi-ML estimators of the conditional mean and variance parameters.

Since the existing simulation evidence indicates that the finite sample size properties of many normality tests could be significantly different from the nominal levels (see e.g. Doornik and Hansen (1994), Jarque and Bera (1987), or White and MacDonald (1980)), the results in Kilian and Demiroglu (2000) suggest that a fruitful avenue for future research would be to consider bootstrap procedures in order to reduce size distortions.

Similarly, given that neither version of our proposed LM test has power against asymmetric alternatives by construction, it would also be worth exploring ways in which they can be complemented with tests for multivariate symmetry. One possibility would be to use the asymmetry component of Mardia's (1970) test for multivariate normality, which is also numerically invariant to the way in which the residuals are orthogonalised. As argued in the previous section, though, if the conditional mean and variance parameters have to be estimated, it would be necessary

to modify his test statistic to make it orthogonal to all the elements of $s_{\mu t}(\mu_0; 0)$ (see Davidson and MacKinnon (1993) for the correction involved in the univariate case). Theorem 1 also suggests an alternative way of testing for multivariate symmetry, which would be based on the fact that the normalised innovations $\frac{\epsilon_t^a(\mu_0)}{\sqrt{\epsilon_t^a(\mu_0)' \epsilon_t^a(\mu_0)}}$ would be uniformly distributed on the unit sphere surface in R^N independently of $\epsilon_t(\mu_0)$ under the null of multivariate normality. For instance, in the case of $N = 1$, the normalised innovations are simply $\frac{\epsilon_t^a(\mu_0)}{|\epsilon_t^a(\mu_0)|} = 2\text{sign}[\epsilon_t^a(\mu_0)] - 1$. But since $\text{sign}[\epsilon_t^a(\mu_0)]$ is distributed as a Bernoulli random variable with parameter $1/2$ independently of $\epsilon_t^a(\mu_0)$, a simple test for univariate symmetry would be the LM test of $H_0 : E\{\text{sign}[\epsilon_t^a(\mu_0)] | \hat{A}_0\} = 1/2$ (see Engle (1984)). Unfortunately, while $\text{sign}[\epsilon_t^a(\mu_0)] - 1/2$ is orthogonal to $\epsilon_t^a(\mu_0) - 1$, it is not orthogonal to $\epsilon_t^a(\mu_0)$, which means that we cannot ignore the fact that μ_0 will often have to be replaced by $\hat{\mu}_T$. Similarly, when $N = 2$, we could decompose $\epsilon_t^a(\mu_0)$ in the polar coordinates $\sqrt{\epsilon_t^a(\mu_0)' \epsilon_t^a(\mu_0)}$ and $\tilde{A}_{1t}(\mu_0) = \arctan[\epsilon_{2t}^a(\mu_0)/\epsilon_{1t}^a(\mu_0)]$, where the angle $\tilde{A}_{1t}(\mu_0)$ should be distributed as a uniform continuous random variable in the interval 0 to 2π . A generalisation of such an approach to higher dimensions could be obtained on the basis of Theorem 2.11 in Fang, Kotz and Ng (1990), who show that the transformation of a spherically symmetric multivariate random vector of dimension $N \geq 2$ to spherical coordinates produces N mutually independent random variables with known distribution.

Appendix

Proofs of results

Proposition 1

We can use Example 2 to write

$$\frac{N'_{00} + 1}{1 - 2'_{00} + '_{00}\&_t(\mu_0)} \pi_t^{\pi}(\mu_0) = \frac{(N'_{00} + 1) \rho_{\>_t + '_{00}(\rho_0 - 2)^{3_t}}}{(1 - 2'_{00}) \rho_{\>_t + '_{00}(\rho_0 - 2)^{3_t}} \rho_{\>_t}} u_t \quad (17)$$

and

$$\frac{N'_{00} + 1}{1 - 2'_{00} + '_{00}\&_t(\mu_0)} \pi_t^{\pi}(\mu_0) \pi_t^{\pi_0}(\mu_0) - I_N = \frac{N'_{00} + 1}{'_{00}} \frac{3_t}{\>_t + 3_t} u_t u_t^0 - I_N \quad (18)$$

The expectation of (17) is clearly zero because all the variables involved are mutually independent, and $E(u_t) = 0$ from Theorem 2.7 in Fang, Kotz and Ng (1990).

The same theorem also implies that $E(u_t u_t^0) = N^{-1} I_N$. In addition, since $3_t = (\>_t + 3_t)$ is an independent beta variate with parameters N and ρ_0 , whose expected value is $N = (\rho_0 + N)$, then (18) will also be 0.

Finally, the vector martingale difference property trivially follows from the fact that u_t , 3_t and $\>_t$ are independent of z_t and I_{t_i-1} by assumption.

Proposition 2

First of all, it is easy to see that:

$$\begin{aligned} E[s_{\cdot t}(\mu_0; 0) | z_t; I_{t_i-1}; \hat{A}_0] &= \frac{N(N+2)}{4} - \frac{(N+2)}{2} E[\&_t(\mu_0) | z_t; I_{t_i-1}; \hat{A}_0] \\ + \frac{1}{4} E[\&_t^2(\mu_0) | z_t; I_{t_i-1}; \hat{A}_0] &= \frac{N(N+2)}{4} - \frac{N(N+2)}{2} + \frac{2N + N^2}{4} = 0 \end{aligned}$$

where we have used the fact that under the null $\&_t(\mu_0)$ is an i.i.d. chi-square variate with N degrees of freedom (see Lemma 1), whose uncentred moment of integer

order r is

$$E(s_t^r) = 2^r \left[r-1 + \frac{N}{2} \right] \left[r-2 + \frac{N}{2} \right] \cdots \left[1 + \frac{N}{2} \right] \frac{N}{2}$$

(see e.g. Mood, Graybill and Boes (1973)).

Similarly, we can show that

$$\begin{aligned} V[s_t(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] &= E[s_t^2(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] - \frac{N(N+2)^2}{4} E[\&_t(\mu_0)|z_t; I_{t-1}; \hat{A}_0]^2 \\ &= \frac{N^2(N+2)^2}{16} - \frac{N(N+2)^2}{4} E[\&_t(\mu_0)|z_t; I_{t-1}; \hat{A}_0]^2 + \frac{(N+2)^2}{5} + \frac{N(N+2)}{8} E[\&_t^2(\mu_0)|z_t; I_{t-1}; \hat{A}_0] \\ &\quad - \frac{N+2}{4} E[\&_t^3(\mu_0)|z_t; I_{t-1}; \hat{A}_0] + \frac{1}{16} E[\&_t^4(\mu_0)|z_t; I_{t-1}; \hat{A}_0] = \frac{N(N+2)}{2} \end{aligned}$$

As for the cross-product terms of the information matrix, note that

$$\begin{aligned} E[s_{\mu t}(\mu_0; 0)s_t(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] &= \frac{\partial^2 \log L(\mu_0)}{\partial \mu \partial \mu} S_t^{1=2}(\mu_0) E[s_t(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] \\ &\quad + \frac{1}{2} \frac{\partial^2 \log L(\mu_0)}{\partial \mu \partial \mu} S_t^{1=2}(\mu_0) \otimes S_t^{1=2}(\mu_0) E[\text{vec}[s_t(\mu_0; 0)s_t(\mu_0; 0)]|z_t; I_{t-1}; \hat{A}_0] \\ &= 0 \end{aligned}$$

If we then use the expressions for the moments up to order six of the spherical multivariate normal distribution (see e.g. Balestra and Holly (1990)), we can show that:

$$\begin{aligned} E[s_t(\mu_0; 0)s_t(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] &= \frac{N(N+2)}{4} E[\&_t^2(\mu_0)|z_t; I_{t-1}; \hat{A}_0] \\ &\quad - \frac{(N+2)}{2} E[\&_t(\mu_0)\&_t(\mu_0)|z_t; I_{t-1}; \hat{A}_0] + \frac{1}{4} E[\&_t^4(\mu_0)|z_t; I_{t-1}; \hat{A}_0] = 0 \end{aligned}$$

and

$$\begin{aligned} E[\&_t^2(\mu_0)\&_t^2(\mu_0)s_t(\mu_0; 0)|z_t; I_{t-1}; \hat{A}_0] &= \frac{N(N+2)}{4} E[\&_t^2(\mu_0)\&_t^2(\mu_0)|z_t; I_{t-1}; \hat{A}_0] \\ &\quad - \frac{N+2}{2} E[\&_t^2(\mu_0)\&_t^2(\mu_0)\&_t(\mu_0)|z_t; I_{t-1}; \hat{A}_0] + \frac{1}{4} E[\&_t^2(\mu_0)\&_t^2(\mu_0)\&_t^2(\mu_0)|z_t; I_{t-1}; \hat{A}_0] \\ &= \frac{N(N+2)}{4} - \frac{(N+2)^2}{2} + \frac{N^2+6N+8}{4} = 0 \end{aligned}$$

as required. Finally, the formula for $V[s_{\mu t}(\mu_0; 0)|\hat{A}_0]$ simply reproduces expression (2.7) in Bollerslev and Wooldridge (1992).

Lemma 1

Under the alternative, the result follows from the fact that

$$\mathfrak{z}_t(\mu_0) = \frac{\mathfrak{y}_t^0 - \mathfrak{y}_t}{\mathfrak{y}_t} = \frac{(\mathfrak{o}_0 - 2) \mathfrak{y}_t^0 - \mathfrak{y}_t}{\mathfrak{y}_t} = \frac{N(\mathfrak{o}_0 - 2)}{\mathfrak{o}_0} \frac{\mathfrak{y}_t^0}{\mathfrak{y}_t}$$

where $\mathfrak{y}_t^0 = \mathfrak{y}_t^0 / \mathfrak{y}_t \sim \hat{A}_N^2$. On this basis, the result for the null follows from the well known fact that \mathfrak{y}_t^0 converges in probability to 1 as $\mathfrak{o}_0 \rightarrow \infty$

Proposition 3

The expressions for $E[s_t(\mu_0; 0)|Z_t; I_{t-1}; \hat{A}_0]$ and $E[s_t^2(\mu_0; 0)|Z_t; I_{t-1}; \hat{A}_0]$ are obtained as in the proof of Proposition 2, except for the fact that under the alternative $\mathfrak{z}_t(\mu_0)$ is proportional to an i.i.d. F variate with N and \mathfrak{o}_0 degrees of freedom (see Lemma 1), whose uncentred moment of integer order $r < \mathfrak{o}_0 = 2$ is

$$E \left[\frac{\mathfrak{y}_t^0}{\mathfrak{y}_t} \right]^r = \frac{\mathfrak{o}_0}{N} \frac{(r-1+N)(r-2+N) \cdots (1+N)(N)}{(-1+\mathfrak{o}_0)(-2+\mathfrak{o}_0) \cdots (-r+1+\mathfrak{o}_0)(-\mathfrak{o}_0)}$$

(see e.g. Mood, Graybill and Boes (1973)). Therefore, the restriction $\mathfrak{o}_0 > 8$ guarantees that the fourth moments of $\mathfrak{z}_t(\mu_0)$ are bounded. Finally, the asymptotic distribution is obtained from a straightforward application of the Lindeberg-Levy central limit theorem for independent and identically distributed observations.

FIGURE 1: Derivative of $\alpha(\eta)$ with respect to η and first order Taylor expansion

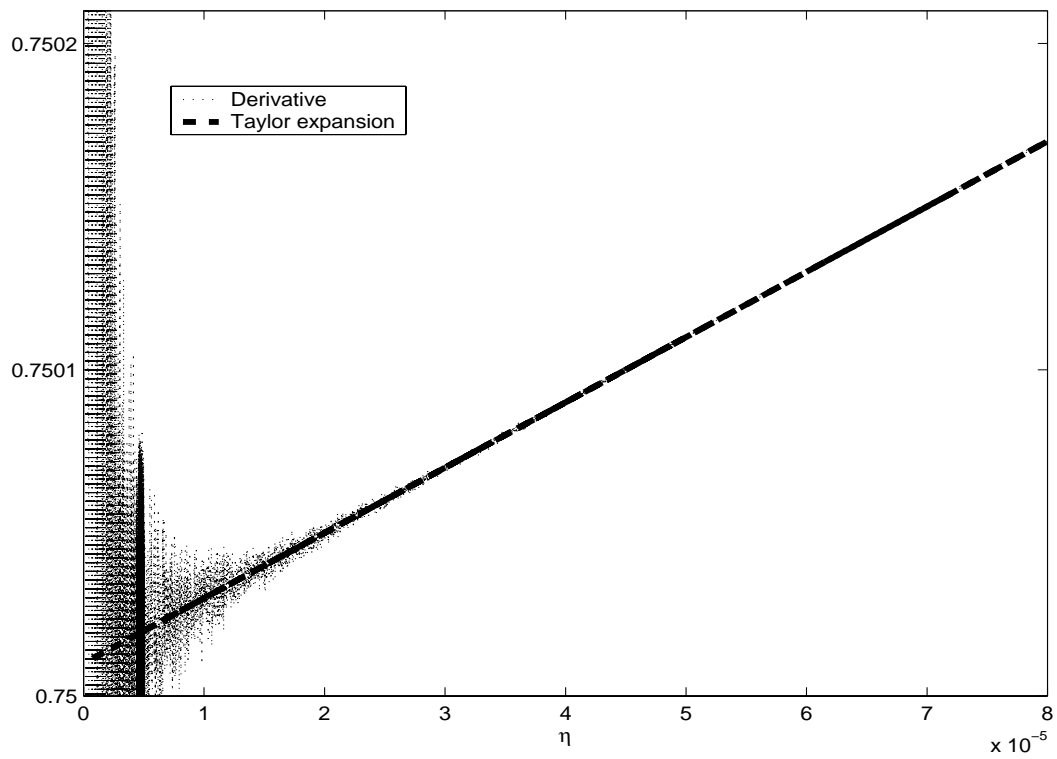


FIGURE 2: Power of the LM test ($T=100, \alpha=5\%$).

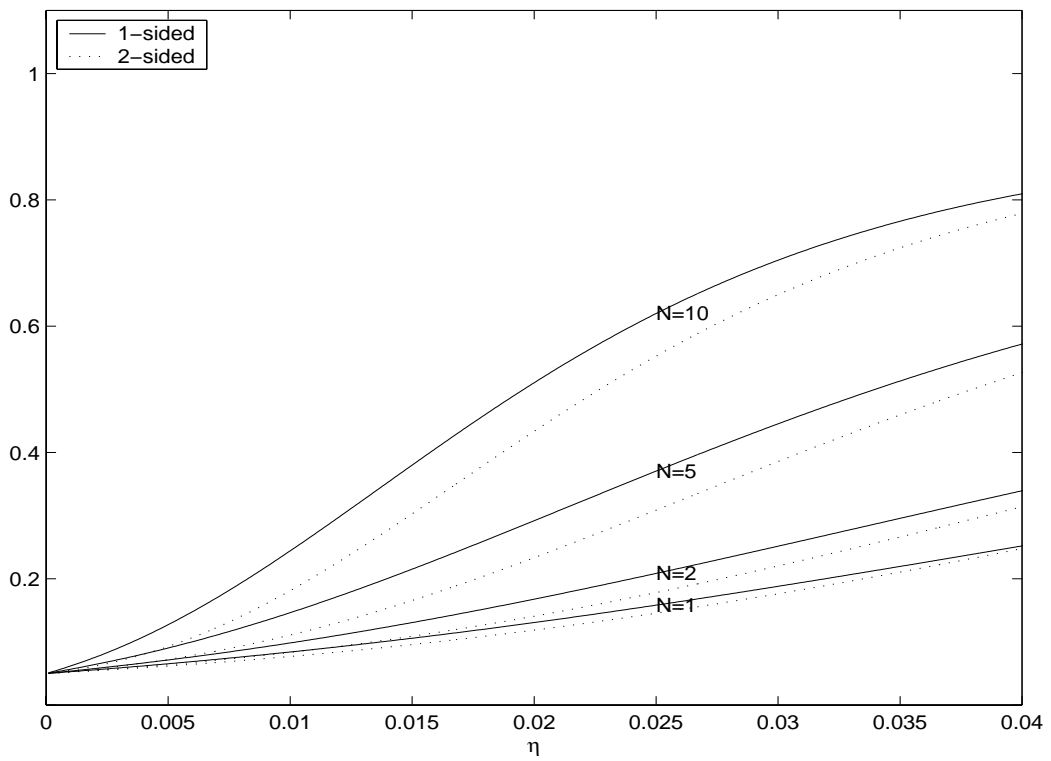


FIGURE 3: Power of the LM test ($T=250, \alpha=5\%$).

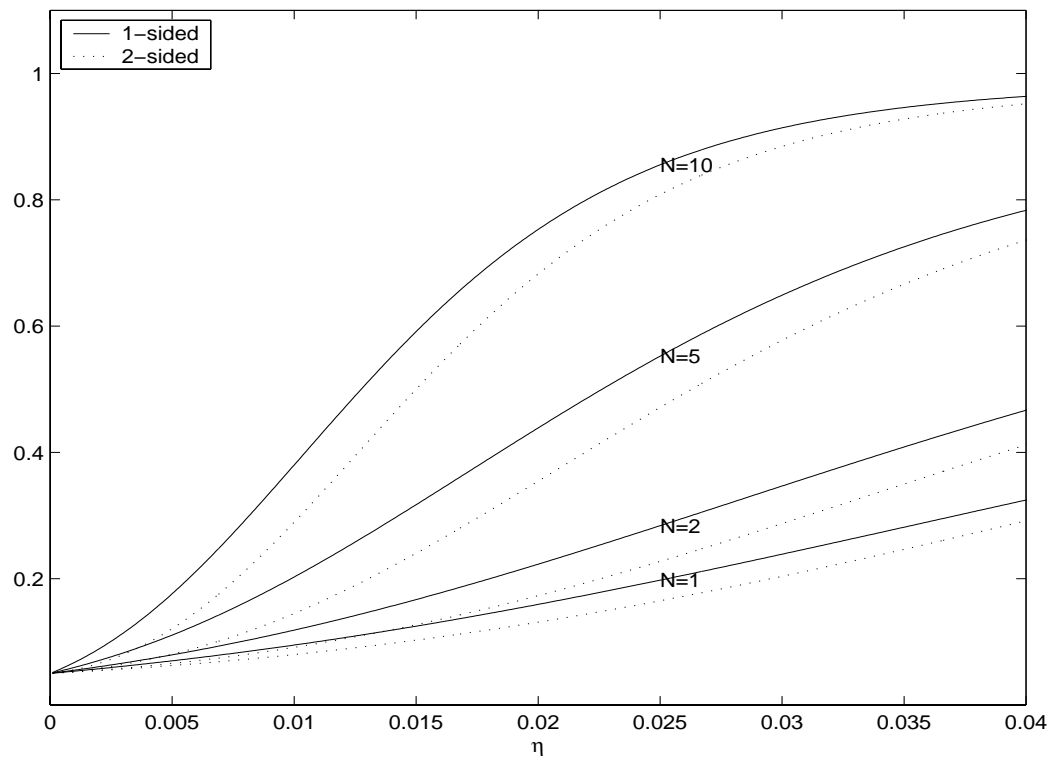
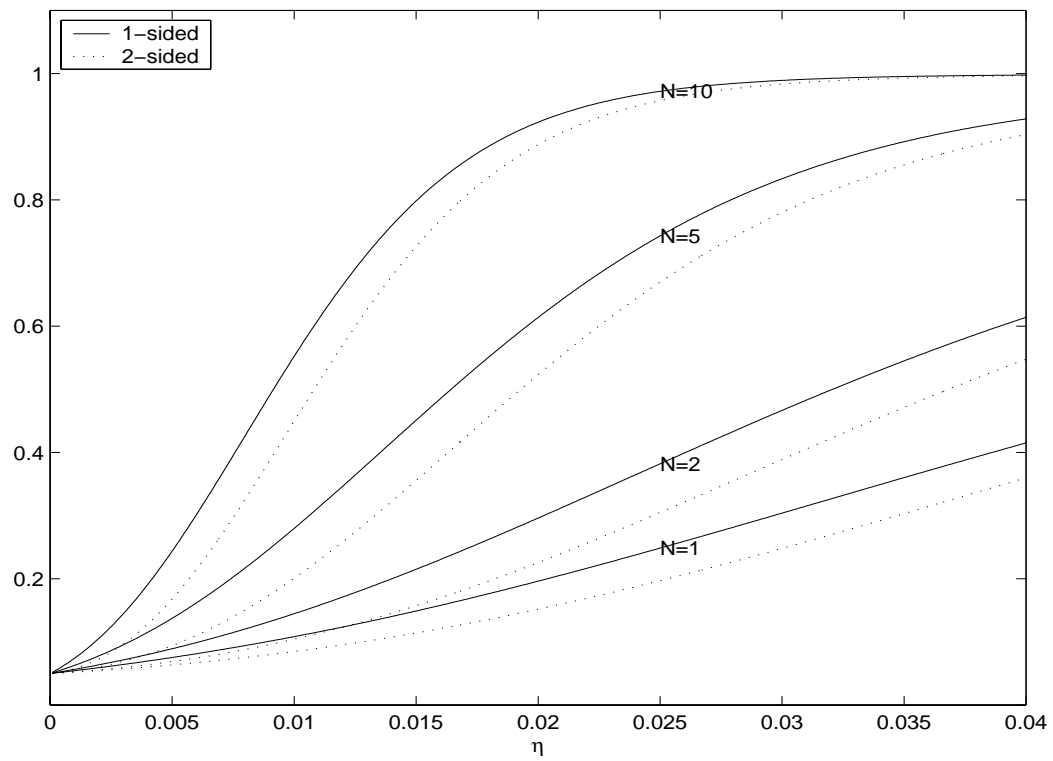


FIGURE 4: Power of the LM test ($T=500, \alpha=5\%$).



References

- Abramowitz, M. and Stegun, I.A. (1964): Handbook of mathematic functions, AMS 55, National Bureau of Standards.
- Andrews, D.W.K. (2000): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica* forthcoming.
- Baillie, R. T., and Bollerslev, T. (1989): "The message in daily exchange rates: a conditional-variance tale", *Journal of Business & Economic Statistics* 7, 297-305.
- Balestra, P. and Holly, A. (1990): "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.
- Bera, A.K. and Jarque, C.M. (1981): "Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence", *Economic Letters*, 7, 313-318.
- Bera, A.K. and McKenzie, C.R. (1987): "Additivity and separability of Lagrange multiplier, likelihood ratio and Wald tests", *Journal of Quantitative Economics* 3, 53-63.
- Bollerslev, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return", *Review of Economics & Statistics* 69, 542-547.
- Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992): "ARCH modeling in finance: a review of the theory and empirical evidence", *Journal of Econometrics* 52, 5-59.
- Bollerslev, T., and J. M. Wooldridge (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.
- Crowder, M.J. (1976): "Maximum likelihood estimation for dependent obser-

vations", *Journal of the Royal Statistical Society B*, 38, 45-53.

Davidson R. and MacKinnon, J.G. (1983): "Small sample properties of alternative forms of the Lagrange multiplier test", *Economics Letters* 12, 269-275.

Davidson R. and MacKinnon, J.G. (1993): *Estimation and inference in econometrics*, Oxford University Press, Oxford.

Demos, A. and Sentana, E. (1998): "Testing for Garch effects: a one-sided approach", *Journal of Econometrics*, 86, 97-127.

Doornik, J.A. and Hansen, H. (1994): "An omnibus test for univariate and multivariate normality", Working Paper W4&91, Nuffield College, Oxford.

Engle, R.F. (1984): "Wald, likelihood ratio and Lagrange multiplier tests in econometrics", in Griliches, Z. and Intriligator, M.D., eds., *Handbook of Econometrics Vol. 2*, 775-826, North Holland.

Fang, K.-T., Kotz, S. and Ng, K.-W. (1990): *Symmetric multivariate and related distributions*, Chapman and Hall.

Fiorentini, G., Calzolari, G. and Panatoni, L. (1996): "Analytical derivatives and the computation of Garch models", *Journal of Applied Econometrics* 11, 399-417.

Gill, P.E., Murray, W. and Wright, M.H. (1981): *Practical optimization*, Academic Press.

Gourieroux C., Holly A. and Monfort A. (1980): "Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Harvard Institute of Economic Research Discussion Paper 770.

Harvey, A., Ruiz, E. and Sentana, E. (1992): "Unobservable component time series models with Arch disturbances", *Journal of Econometrics* 52, 129-158.

Kiefer, N.M. and Salmon, M. (1983): "Testing normality in econometric models", *Economic Letters* 11, 123-127.

Kilian, L. and Demiroglu, U. (2000): "Residual-based test for normality in

autoregressions: asymptotic theory and simulation evidence", *Journal of Business & Economic Statistics* 18, 40-50.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economic Letters*, 6, 255-259.

Jarque, C.M. and Bera, A.K. (1987): "A test for normality of observations and regression residuals", *International Statistical Review* 55, 163-172.

Lee, J.H.H. and King, M.L. (1993): "A locally most mean powerful based score test for Arch and Garch regression disturbances", *Journal of Business & Economic Statistics* 11, 17-27.

Lütkepohl, H. (1993): *Introduction to multiple time series analysis*, 2nd ed., Springer-Verlag.

Mardia, K.V. (1970): "Measures of multivariate skewness and kurtosis with applications", *Biometrika* 57, 519-530.

McCullough, B. D., and Vinod, H.D. (1999): "The numerical reliability of econometric software", *Journal of Economic Literature* 37, 633-665.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974): *Introduction to the theory of Statistics*, 3rd edition, McGraw Hill.

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-99.

Pesaran, M.H. and Pesaran, B. (1997): *Working with Micro...t 4.0: interactive econometric analysis*, Oxford University Press.

Sentana, E. (1991): "Quadratic Arch models: a potential re-interpretation of Arch models", *LSE Financial Markets Group Discussion Paper* 122.

Sentana, E. (1995): "Quadratic Arch models", *Review of Economic Studies* 62, 639-661.

Sentana, E. (2000): "Factor representing portfolios in large asset markets", CEMFI Working Paper 0001.

White, H. and MacDonald, G.M. (1980): "Some large sample test for non-normality in the linear regression model", Journal of the American Statistical Association 75, 16-28.

Zellner, A. (1971): An introduction to Bayesian inference in econometrics, Wiley.